INFORMATION AND CONTROL 16, 203–224 (1970)

Two-State Two-Symbol Probabilistic Automata

TOSHIO YASUI*

Department of Electronics, Faculty of Engineering, Kyoto University, Kyoto, Japan

AND

SHUZO YAJIMA

Department of Electrical Engineering II, Faculty of Engineering, Kyoto University, Kyoto, Japan

In this paper, we discuss the algebraic treatments of probabilistic automata with two states. We derive the result that the matrix product corresponding to a given input tape can be decomposed into the sum of a finite number of fundamental matrices which are determined by the matrices corresponding to the input symbols. We introduced the new concept of *probabilistic automata completely isolated by the* L-th *approximation*. With respect to these automata, all tapes of length greater than or equal to $L + 1$ can be classified into 2^{L+1} sets by means of their $(L + 1)$-suffixes. By using this concept, we demonstrate that a two-input symbol actual automaton with two states can realize any definite event.

I. INTRODUCTION

On developing the theory of probabilistic automata, a special class of probabilistic automata has played an important role. Rabin [5] used a two-state probabilistic automaton in order to show that the class of events realized by probabilistic automata is strictly larger than the class of events realized by deterministic automata.

The precise definition and some properties of m-adic probabilistic automata, originating from this Rabin's idea, were given by Paz [3]. Recently, Salomaa [6] also discussed m-adic probabilistic automata and m-adic

* From September, 1968, T. Yasui is with the Department of Computer Science, University of Illinois, Urbana, Illinois 61801.

643/16/3-1

events. This class of probabilistic automata has a special feature; there is a one to one correspondence between the set of all inputs and their corresponding probabilities in the interval [0, 1]. Namely, when the output probability $p(x)$ of the probabilistic automaton is written in the scale of m adic, the sequence of coefficients indicates the input tape x. In general, however, for an input tape x, it is not easy to find the output probability $p(x)$ because we must calculate matrix products corresponding to the tape x.

The object of this paper is to introduce mathematical treatments of probabilistic automata with two states and to present algebraic considerations on them from the standpoint of characteristic values of stochastic matrices. Although algebraic treatments of Markov chains, in which powers of a single stochastic matrix can be easily obtained are well-known (e.g., Feller [1]), it is rather difficult to apply them to probabilistic automata directly because the treatment involving products of several distinct stochastic matrices has not been explored. Products involving 2×2 stochastic matrices, however, can be treated by substituting for each matrix in the product its decomposed expression, that is, a sum of fundamental matrices. We then show that the resulting product can also be written as the sum of a finite number of fundamental matrices.

After considering several properties of a set of stochastic matrices of order two (Section III) and introducing the approximation method (Section IV), a new concept of a *completely isolated set of stochastic matrices* is introduced (Section V).

The set of output probabilities $p(x)$ of a completely isolated probabilistic automaton is nondense in the interval [0, 1] and this automaton exhibits a behavior similar to nondense m-adic probabilistic automata. 3-adic nondense probabilistic automata in Rabin [5] are special cases of a completely isolated automata. In Section VI, by using a definition of event which is slightly different from Rabin's, we are led to the conclusion that there exist two-state two-symbol actual automata which realize any definite event.

II. Basic Definitions

In this section, basic notation and definitions are given by referring to Rabin [5].

Let Σ be a finite nonempty *alphabet*. The elements of Σ are called *symbols*. The set of all finite sequences of elements of Σ will be denoted by Σ^* and the elements of Σ^* will be called *tapes*. The empty tape will denoted by $\Lambda(\Lambda \in \Sigma^*)$. Subsets of Σ^* will be called sometimes *events*. Letter σ will usually denote an element of Σ and letters x, y, z will always denote tapes. If x and y are tapes, then xy denotes the concatenation of x and y. For any tape $x = \sigma_1\sigma_2 \cdots \sigma_k$, the length $l(x)$ of x is k. The tape z is said to be a *k suffix* of x if $x = yz$ for some tape y and $l(z) = k$. Also, the tape y is said to be a *k prefix* of x if $x = yz$ for some tape z and $l(y) = k$.

DEFINITION 1. A probabilistic automaton PA over the alphabet Σ is a system $PA = (S, M, s_1, F)$, where $S = (s_1, s_2, \ldots, s_n)$ is a finite set of states, M is a function from $S \times \Sigma$ into $[0, 1]^n$ such that for $(s, \sigma) \in S \times \Sigma$

$$M(s, \sigma) = [p_1(s, \sigma), p_2(s, \sigma), \ldots, p_n(s, \sigma)],$$
$$0 \leqslant p_i(s, \sigma), \qquad \sum_i p_i(s, \sigma) = 1,$$

where $p_i(s, \sigma)$ is the transition probability of PA moving from the state s to the state s_i when the input symbol is σ, $s_1 \in S$ (the *initial state*), and $F \subseteq S$ (the set of *designated final states*).

DEFINITION 2. For $\sigma \in \Sigma$ and $x = \sigma_1\sigma_2 \cdots \sigma_m$, define the $n \times n$ matrices $A(\sigma)$ and $A(x)$ by

$$A(\sigma) = [p_j(s_i, \sigma)]$$
$$A(x) = A(\sigma_1)\, A(\sigma_2) \cdots A(\sigma_m) = [p_j(s_i, x)] \qquad \text{for} \quad i, j = 1, 2, \ldots, n.$$

A simple calculation (involving induction on m) will show that the (i, j)-th element $p_j(s_i, x)$ of the matrix $A(x)$ is the probability of PA moving from state s_i to state s_j when, the input tape is x. $A(\sigma)$ and $A(x)$ are *stochastic*[1] and are called the matrix of the input symbol σ and the matrix of the input tape x.

In the rest of the paper, we use the set of matrices $A(\sigma)$ instead of the table M of the set of transition probabilities.

DEFINITION 3. Let $PA = (S, M, s_1, F)$ and $F = (s_{i_1}, \ldots, s_{i_r})$, $I = (i_1, \ldots, i_r)$. Let the output probability $p(x)$ be defined as,

$$p(x) = \sum_{i \in I} p_i(s_1, x).$$

$p(x)$ is clearly the probability for PA when started in s_1, to enter a state which is a member of F by the input tape x.

[1] An n by n stochastic matrix is defined to be any $n \times n$ real matrix $A = [a_{ij}]$ which satisfies the two properties $0 \leqslant a_{ij} \leqslant 1$, $\Sigma_j a_{ij} = 1$ for $i = 1, 2, \ldots, n$.

DEFINITION 4. Let PA be a probabilistic automaton, c be a real number $0 \leqslant c \leqslant 1$, and $\epsilon > 0$ be an arbitrary small number. Then a set of tapes $T(PA, c, \epsilon)$ is defined by

$$T(PA, c, \epsilon) = \{x \mid x \in \Sigma^*, \mid p(x) - c \mid < \epsilon\}.$$

If $x \in T(PA, c, \epsilon)$, we say that x is accepted by PA with cut point c and error ϵ.

This definition shows that the set of tapes, under which the probability of reaching from an initial state s_1 to one of final states is between $c - \epsilon$ and $c + \epsilon$, is called an event realized by PA with c and ϵ.

Remark. The above Definition 4 is slightly different from Rabin's Definition 7.[2] However, we can show that the new definition leads to the same family of events as Rabin's original definition if we allow the number of states of a probabilistic automaton to increase. The sketch of the proof can be given as follows. Assume that T is an event accepted by a probabilistic automaton according to Rabin's original definition. Without loss of generality, we can assume that T is accepted by a probabilistic automaton PA with cut point $\frac{1}{2}$,

$$T(PA, \tfrac{1}{2}) = \{x \mid x \in \Sigma^*, p(x) > \tfrac{1}{2}\}.$$

We may write this in the form

$$T(PA, 1, \tfrac{1}{2}) = \{x \mid x \in \Sigma^*, \mid p(x) - 1 \mid < \tfrac{1}{2}\},$$

which shows that T is acceptable according to the new definition. Conversely, assume that there is a set of tapes T such as

$$\begin{aligned} T(PA, c, \epsilon) &= \{x \mid x \in \Sigma^*, \mid p(x) - c \mid < \epsilon\} \\ &= \{x \mid x \in \Sigma^*, (p(x) - c)^2 < \epsilon^2\}. \end{aligned}$$

Using the technique of Kronecker products, it can be shown that the right side is an event acceptable in the original Rabin's sense (Turakainen [7]).

[2] Let $T(PA, c)$ be defined by $T(PA, c) = \{x \mid x \in \Sigma^*, p(x) > c\}$. If $x \in T(PA, c)$, then x is said to be accepted by PA with cut point c.

III. Properties of Sets of 2 × 2 Stochastic Matrices

In this section, we discuss algebraic properties of sets of 2×2 stochastic matrices. Let A and B be the following two stochastic matrices, corresponding to symbols 0 and 1, respectively, (See Fig. 1.),

$$A = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}, \qquad B = \begin{bmatrix} 1-c & c \\ d & 1-d \end{bmatrix},$$

where $0 \leqslant a, b, c, d \leqslant 1$.

The *characteristic roots* of the matrix A are $\lambda_1 = 1$, $\lambda_2 = (1 - a - b)$. We can find the i-th *characteristic column vector* $\mathbf{x}_i$, and the i-th *characteristic row vector* $\mathbf{y}_i$, for $i = 1, 2$, as follows:

$$\mathbf{X}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \mathbf{X}_2 = \begin{bmatrix} a/(a+b) \\ -b/(a+b) \end{bmatrix},$$

$$\mathbf{Y}_1 = [b/(a+b) \;\; a/(a+b)], \qquad \mathbf{Y}_2 = [1 \;\; -1],$$

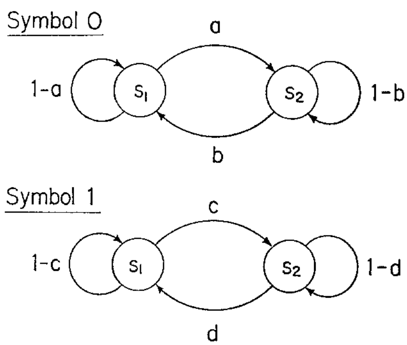

Fig. 1. Transition probabilities of two-state two-symbol probabilistic automaton.

These characteristic vectors $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{y}_1$, and $\mathbf{y}_2$ can be chosen such that

$$PAP^{-1} = Z, \tag{1}$$

where

$$P = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}, \qquad P^{-1} = [\mathbf{x}_1 \;\; \mathbf{x}_2], \qquad \text{and} \qquad Z = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$

Then, because $PP^{-1} = E$ (E: *unit matrix*), for $i, j = 1, 2$,

$\mathbf{y}_i \mathbf{x}_j = \delta_{ij}$, where *Kroneker delta* $\delta_{ij} = 1$ if $i = j$ and 0, otherwise. (2)

Let the i-th *fundamental matrix* A_i of A, for $i = 1, 2$, be defined by

$$\begin{aligned} A_1 = \mathbf{x}_1\mathbf{y}_1 &= \begin{bmatrix} b/(a+b) & a/(a+b) \\ b/(a+b) & a/(a+b) \end{bmatrix}, \\ A_2 = \mathbf{x}_2\mathbf{y}_2 &= \begin{bmatrix} a/(a+b) & -a/(a+b) \\ -b/(a+b) & b/(a+b) \end{bmatrix}. \end{aligned} \tag{3a}$$

Then, by using (1),

$$A = P^{-1}ZP = \lambda_1\mathbf{x}_1\mathbf{y}_1 + \lambda_2\mathbf{x}_2\mathbf{y}_2 ,$$

that is,

$$A = A_1 + (1 - a - b)\, A_2$$

can be obtained. As for B, we can obtain

$$B = B_1 + (1 - c - d)\, B_2 ,$$

where

$$\begin{aligned} B_1 = \mathbf{u}_1\mathbf{v}_1 &= \begin{bmatrix} d/(c+d) & c/(c+d) \\ d/(c+d) & c/(c+d) \end{bmatrix}, \\ B_2 = \mathbf{u}_2\mathbf{v}_2 &= \begin{bmatrix} c/(c+d) & -c/(c+d) \\ -d/(c+d) & d/(c+d) \end{bmatrix}, \end{aligned} \tag{3b}$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the i-th characteristic column and row vector of B, respectively.

Next, we express the given stochastic matrices of the input symbol 0 and 1 in (3) and calculate the matrices of input tapes, by the fundamental matrices. It will be shown that not only $A, A^2, \ldots, B, B^2, \ldots,$ but also their products AB, BA, AAB, etc. can be effectively represented by using these characteristic values and fundamental matrices. Now we present some properties of products of the fundamental matrices A_i and B_i ($i = 1, 2$).

LEMMA 1. *The following relations are satisfied:*

(1) $A_iA_j = \delta_{ij}A_j$

(2) $A_iB_1 = \delta_{i1}B_1$

(3) $A_1B_2 = -B_1A_2$

(4) $A_2B_2 = A_2$.

Proof. (1) From the fact that $A_iA_j = \mathbf{x}_i\mathbf{y}_i\mathbf{x}_j\mathbf{y}_j$ and the scalar product $\mathbf{y}_i\mathbf{x}_j = \delta_{ij}$, $A_iA_j = \delta_{ij}\mathbf{x}_i\mathbf{y}_j = \delta_{ij}A_j$ holds.

(2) According to (3a) and (3b),

$$A_1B_1 = \mathbf{x}_1\mathbf{y}_1\mathbf{u}_1\mathbf{v}_1 \quad \text{and} \quad A_2B_1 = \mathbf{x}_2\mathbf{y}_2\mathbf{u}_1\mathbf{v}_1 .$$

By using $\mathbf{x}_1 = \mathbf{u}_1$ and $\mathbf{y}_i\mathbf{x}_j = \delta_{ij}$, we obtain

$$A_1B_1 = (\mathbf{y}_1\mathbf{u}_1)\,\mathbf{x}_1\mathbf{v}_1 = \mathbf{x}_1\mathbf{v}_1 = B_1 \quad \text{and} \quad A_2B_1 = (\mathbf{y}_2\mathbf{u}_1)\,\mathbf{x}_2\mathbf{v}_1 = [0].$$

(3) In general,

$$A_1 + A_2 = E \quad \text{and} \quad B_1 + B_2 = E$$

are true. Therefore, it follows that

$$A_1B_2 = (E - A_1)(E - B_1) = E - A_2 - B_1 = A_1 - B_1$$
$$B_1A_2 = (E - B_2)(E - A_1) = E - B_2 - A_1 = B_1 - A_1 ,$$

so that

$$A_1B_2 = -B_1A_2 .$$

(4) We also obtain

$$A_2B_2 = A_2(E - B_1) = A_2 ,$$

because $A_2B_1 = [0]$ from (2). Q.E.D.

We define another *fundamental matrix* H as follows,

$$H = A_1B_2 = -B_1A_2 = \frac{bc - ad}{(a + b)(c + d)} \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}.$$

Table I is a multiplication table which is convenient to use to calculate the matrix products made from A and B. For example, the product A_1B_2 is written at the intersection of A_1's row and B_2's column, and the notation 0 denotes a *null matrix*.

PROPOSITION 1. *The relations shown in Table* I *hold.*

The above proposition can be easily shown by using Lemma 1.

Table I shows that the set $\{A_1 , A_2 , B_1 , B_2 , H, \text{null matrix}\}$ is closed under matrix multiplications. In the case of the set of $n \times n$ stochastic matrices ($n \geqslant 3$), this closure property does not hold. However, in the special case of the set of k 2×2 stochastic matrices, the following theorem is obtained.

TABLE I

Multiplication Table of the Set of Two Stochastic Matrices A and B

	A_1	A_2	B_1	B_2	H
A_1	A_1	0	B_1	H	H
A_2	0	A_2	0	A_2	0
B_1	A_1	$-H$	B_1	0	H
B_2	0	B_2	0	B_2	0
H	0	H	0	H	0

THEOREM 1. *For each set of* k 2×2 *stochastic matrices, there exists a set of* $\binom{k}{2} + 2k + 1$ 2×2 *matrices that is closed under matrix multiplication.*

Proof. The argument used for the case $k = 2$ can be extended to this general case. Note that $\binom{k}{2}$ matrices are produced to preserve the closure property instead of the one matrix H. Q.E.D.

Now we present the characteristic expansion form of the matrix of the input tape x. We express the matrix $A(x)$ of the tape x of length m as $C^{(1)}C^{(2)} \cdots C^{(m)}$, where $C^{(j)}$ represents $A(B)$, where the j-th symbol of a given tape is 0(1). Furthermore, we let the i-th fundamental matrix and characteristic value of $C^{(j)}$ be represented by $C_i^{(j)}$ and $v_i^{(j)}$, respectively.

THEOREM 2. *A product of the stochastic matrices,* $C^{(1)}C^{(2)} \cdots C^{(m)}$ *can be represented as follows:*

$$\begin{aligned} C^{(1)}C^{(2)} \cdots C^{(m)} &= \prod_{j=1}^{m} C^{(j)} = \prod_{j=1}^{m} \sum_{i=1}^{2} (v_i^{(j)} C_i^{(j)}) \\ &= C_1^{(m)} + v_2^{(m)} C_1^{(m-1)} C_2^{(m)} + v_2^{(m-1)} v_2^{(m)} C_1^{(m-2)} C_2^{(m-1)} + \cdots \\ &\quad + v_2^{(2)} v_2^{(3)} \cdots v_2^{(m)} C_1^{(1)} C_2^{(2)} + v_2^{(1)} v_2^{(2)} \cdots v_2^{(m)} C_2^{(1)} \\ &= C_1^{(m)} + \sum_{k=1}^{m} \left(\prod_{j=k}^{m} v_2^{(j)} \right) C_1^{(k-1)} C_2^{(k)}, \end{aligned} \tag{4}$$

where $C_1^{(0)} = E$ and the notation $\prod_{j=1}^{m}$ implies matrix multiplication from the right, by the order $j = 1, 2,\ldots, m$.

Remark. Note that for some a and b,

$$C^{(1)}C^{(2)} \cdots C^{(m)} = C_1^{(m)} + aH + bC_2^{(1)}.$$

(3.4) is said the *characteristic expansion form of* $C^{(1)}C^{(2)} \cdots C^{(m)}$.

Proof. From the fact $v_1^{(j)} = 1$ for all j, we can rewrite $C^{(j)}$ as follows,

$$C^{(j)} = C_1^{(j)} + v_2^{(j)}C_2^{(j)}$$

for $j = 1, 2, \ldots, m$.

From Proposition 1, $C_i^{(j)}C_1^{(k)} = \delta_{i1}C_1^{(k)}$ and $C_2^{(j)}C_2^{(k)} = C_2^{(j)}$ hold. Therefore, we can write the product $C^{(j)}C^{(k)}$ in the following form,

$$C^{(j)}C^{(k)} = C_1^{(k)} + v_2^{(k)}C_1^{(j)}C_2^{(k)} + v_2^{(j)}v_2^{(k)}C_2^{(j)}.$$

The expansion (3.4) can be derived by the iteration of this process. Q.E.D.

Unless λ_2 and/or μ_2 is the value 1 or -1, then $| v_2^{(j)} | < 1$ holds. Thus, the coefficient $b = v_2^{(1)}v_2^{(2)} \cdots v_2^{(m)}$ goes to value 0 as the length m becomes infinitely large. Note that each matrix of $C_1^{(i)}$ and H has identical rows so that $C^{(1)}C^{(2)} \cdots C^{(m)}$ converges to a matrix having identical rows. Therefore, we are led to Corollary 3, obtained graphically by Paz [2].

COROLLARY 3. *If no matrix corresponding to the input symbol* 0 *or* 1 *is either* I *or* J, *where*

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad J = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

then it has a quasidefinite table.[3]

Proof. A 2×2 stochastic matrix, which is neither I nor J, has the two characteristic values 1 and some number less than 1. A similar result can be shown for the case of multisymbol by referring to Theorem 1. Q.E.D.

EXAMPLE 1. We give $C^{(1)}C^{(2)}C^{(3)}C^{(4)} = ABAA$ as an example. Then $C^{(1)} = C^{(3)} = C^{(4)} = A$ and $C^{(2)} = B$, $C_i^{(1)} = C_i^{(3)} = C_i^{(4)} = A_i$ and $C_i^{(2)} = B_i$, and $v_2^{(1)} = v_2^{(3)} = v_2^{(4)} = \lambda_2$ and $v_2^{(2)} = \mu_2$.

[3] A *quasidefinite table* is defined by $[S, A(\sigma)\text{'s}]$, where S is a set of states and $A(\sigma)$'s satisfy the following condition:
for any $\epsilon > 0$, there exists such an integer k that for any tape x of length no less than k, $|||A(X)||| < \epsilon$ where for a matrix $A = [a_{ij}]$, $|||A|||$ means $|||A||| = \max_j \max_{i_1, i_2} | a_{i_1 j} - a_{i_2 j} |$

Applying Theorem 2 to this example and considering $A_i A_j = \delta_{ij} A_j$, we obtain the following characteristic expansion form of $ABAA$.

$$\begin{aligned} ABAA &= C_1^{(4)} + v_2^{(4)} C_1^{(3)} C_2^{(4)} + v_2^{(3)} v_2^{(4)} C_1^{(2)} C_2^{(3)} + v_2^{(2)} v_2^{(3)} v_2^{(4)} C_1^{(1)} C_2^{(2)} \\ &\quad + v_2^{(1)} v_2^{(2)} v_2^{(3)} v_2^{(4)} C_2^{(2)} \\ &= A_1 + \lambda_2 A_1 A_2 + \lambda_2{}^2 B_1 A_2 + \mu_2 \lambda_2{}^2 A_1 B_2 + \mu_2 \lambda_2{}^3 A_2 \\ &= A_1 + (-\lambda_2{}^2 + \mu_2 \lambda_2{}^2) H + \mu_2 \lambda_2{}^3 A_2 . \end{aligned}$$

In this example, the coefficients a and b are $(-\lambda_2{}^2 + \mu_2\lambda_2{}^2)$ and $\mu_2\lambda_2{}^3$, respectively.

IV. Approximation to Products of Stochastic Matrices

In the previous section, we obtained the characteristic expansion form of the matrix of the input tape x. Now we try to approximate its product. Corresponding to the matrix $A(x)$ of the input tape x of length more than L, we will define the following stochastic matrix $\bar{A}^{(\mathrm{L})}(x)$, which is an approximation to $A(x)$.

Definition 5. The L-th *approximate matrix* $\bar{A}^{(\mathrm{L})}(x)$ is defined to be the matrix obtained by eliminating from the expression (4) of $A(x)$ those terms which have more than or equal to L $v_2^{(j)}$'s as factors. Thus, $\bar{A}^{(\mathrm{L})}(x)$ can be written as follows by Theorem 2:

$$\bar{A}^{(\mathrm{L})}(x) = C_1^{(m)} + v_2^{(m)} C_1^{(m-1)} C_2^{(m)} + \cdots + v_2^{(m-\mathrm{L}+1)} \cdots v_2^{(m)} C_1^{(m-\mathrm{L})} C_2^{(m-\mathrm{L}+1)}.$$

Proposition 2. *Let $\Sigma^* z$ designate any tape that has suffix z, then for any tape z of length* L + 1, *the following relation holds:*

$$\bar{A}^{(\mathrm{L})}(z) = \bar{A}^{(\mathrm{L})}(\Sigma^* z).$$

Proof. From Definition 5, the matrix $\bar{A}^{(\mathrm{L})}(\Sigma^* z)$ can be represented by only the (L + 1)-suffix z of $\Sigma^* z$. Q.E.D.

In Table II, we illustrate the L-th approximate matrix to all tapes of length (L + 1) for L = 0, 1, 2. It can be shown that every approximate matrix has identical rows. Therefore, for each L, we can draw a diagram of the set of L-th approximate matrices as in Fig. 2. This is similar to the state transition diagram of a (L + 1)-definite finite automaton. In Fig. 2 of case L = 2, for example, $\bar{A}^{(2)}(x)$ to the tape $x = 110$, is changed to $\bar{A}^{(2)}(101)$ under input symbol 1 because $\bar{A}^{(2)}(1101) = \bar{A}^{(2)}(101)$.

TABLE II

The L-th Approximate Matrices for $L = 0, 1, 2$

(1) $L = 0$	(3) $L = 2$
$\bar{A}^{(0)} = A_1$	$\bar{A}^{(2)}(000) = A_1$
$\bar{A}^{(0)}(1) = B_1$	$\bar{A}^{(2)}(100) = A_1 - \lambda^2 H$
(2) $L = 1$	$\bar{A}^{(2)}(010) = A_1 - \lambda H + \lambda\mu H$
	$\bar{A}^{(2)}(110) = A_1 - \lambda H$
$\bar{A}^{(1)}(00) = A_1$	$\bar{A}^{(2)}(001) = B_1 + \mu H$
$\bar{A}^{(1)}(10) = A_1 - \lambda H$	$\bar{A}^{(2)}(101) = B_1 + \mu H - \lambda\mu H$
$\bar{A}^{(1)}(01) = B_1 + \mu H$	$\bar{A}^{(2)}(011) = B_1 + \mu^2 H$
$\bar{A}^{(1)}(11) = B_1$	$\bar{A}^{(2)}(111) = B_1$

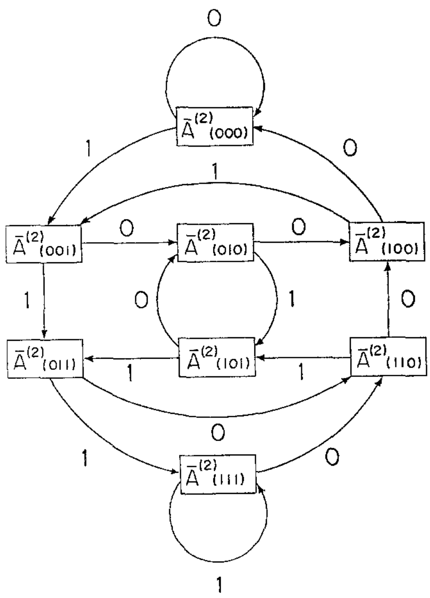

FIG. 2. Transition diagram of the second approximate matrices $\bar{A}^{(2)}(x)$'s.

DEFINITION 6. Let z be any tape of length L + 1 and let y be any tape, then an *error by the* L-th *approximation* $\epsilon^{(L)}$ is defined as,

$$\epsilon^{(L)} = \max_{yz} \| A(yz) - \bar{A}^{(L)}(z) \|,$$

where $\| D \|$ represents the absolute value of a designated element of the matrix D.

THEOREM 4. *We define* δ *as* $\delta = \max\{|\lambda_2|, |\mu_2|\}$. *If* $\delta < 1$, *then*

$$\epsilon^{(L)} < \delta^{L+1} \frac{1}{1-\delta} \max\{\|H\|, \|A_2\|, \|B_2\|\} \qquad (\equiv \bar{\epsilon}^{(L)}).$$

We can define a *maximum error of the* L-th *approximation* $\bar{\epsilon}^{(L)}$ as the right side of the above inequality. $\bar{\epsilon}^{(L)}$ depends only on the matrices A and B of input symbols 0 and 1, and L.

Proof. We assume that yz is a tape of length m, and z is its $(L+1)$-suffix, then from Theorem 2 and Definition 5 we obtain

$$\begin{aligned} A(yz) &= C^{(1)}C^{(2)} \cdots C^{(m)} \\ &= \bar{A}^{(L)}(z) + v_2^{(m-L)} \cdots v_2^{(m)} C_1^{(m-L-1)} C_2^{(m-L)} + \cdots \\ &\quad + v_2^{(2)} \cdots v_2^{(m)} C_1^{(1)} C_2^{(2)} + v_2^{(1)} \cdots v_2^{(m)} C_2^{(1)}. \end{aligned}$$

According to Proposition 1, we know that

$$C_1^{(j)} C_2^{(k)} = H \qquad \text{or} \qquad -H,$$

therefore, it follows that

$$\begin{aligned} \|A(yz) - \bar{A}^{(L)}(z)\| &= |v_2^{(m-L)} \cdots v_2^{(m)}| \, \|H\| + \cdots + |v_2^{(2)} \cdots v_2^{(m)}| \, \|H\| \\ &\quad + |v_2^{(1)} \cdots v_2^{(m)}| \, \|C_2^{(1)}\| \\ &= (\delta^{L+1} + \cdots + \delta^{m-1}) \|H\| + \delta^m \|C_2^{(1)}\| \\ &= (\delta^{L+1} + \cdots + \delta^{m-1} + \delta^m) \max\{\|H\|, \|A_2\|, \|B_2\|\}. \end{aligned}$$

Since $\delta < 1$, the left side converges to

$$\delta^{L+1} \frac{1}{1-\delta} \max\{\|H\|, \|A_2\|, \|B_2\|\} \qquad (\equiv \bar{\epsilon}^{L})$$

as m goes to infinity. Q.E.D.

The above theorem implies that when the maximum error of the L-th approximation $\bar{\epsilon}^{L}$ of a given probabilistic automaton is smaller, then $\bar{A}^{(L)}(\Sigma^* z)$ is closer to $A(z)$. In the following discussion, we treat a probabilistic automaton for which $\delta < 1$ holds.

Now we proceed with more detailed arguments about the L-th approximation.

THEOREM 5. *Let $z = \sigma z'$ be the tape of length* $\mathrm{L} + 1$, *where σ and z' are the* 1-*prefix and* L-*suffix of the tape z, respectively, then, for any integer $k(\geqslant 1)$,*

$$\bar{A}^{(\mathrm{L})}(z) = \bar{A}^{(\mathrm{L}+k)}(\sigma^k z)$$

holds. Namely, the L-th *approximate matrix to the tape of length* $\mathrm{L} + 1$ *equals the* $(\mathrm{L} + k)$-th *approximate matrix to the tape $\sigma^k z$ of length* $\mathrm{L} + k + 1$, *where σ is* 1-*prefix of tape z.*

Proof. Let $A(\sigma^k z)$ be represented by $C^{(1)}C^{(2)} \cdots C^{(\mathrm{L}+k+1)}$. By referring to Definition 5, we obtain the following relations:

$$\begin{aligned}
\bar{A}^{(\mathrm{L}+k)}(\sigma^k z) &= C_1^{(\mathrm{L}+k+1)} + v_2^{(\mathrm{L}+k+1)} C_1^{(\mathrm{L}+k)} C_2^{(\mathrm{L}+k+1)} + \cdots \\
&\quad + v_2^{(k+2)} \cdots v_2^{(\mathrm{L}+k+1)} C_1^{(k+1)} C_2^{(k+2)} + \cdots + v_2^{(1)} \cdots v_2^{(\mathrm{L}+k+1)} C_1^{(1)} C_2^{(2)} \\
&= \bar{A}^{(\mathrm{L})}(\sigma^k z) + v_2^{(k+1)} \cdots v_2^{(\mathrm{L}+k+1)} C_1^{(k)} C_2^{(k+1)} + \cdots \\
&\quad + v_2^{(1)} \cdots v_2^{(\mathrm{L}+k+1)} C_1^{(1)} C_2^{(2)}.
\end{aligned}$$

From Proposition 2,

$$\bar{A}^{(\mathrm{L})}(\sigma^k z) = \bar{A}^{(\mathrm{L})}(z)$$

holds. Since the tape $\sigma^k z$ is of length $\mathrm{L} + k + 1$ and it has σ^{k+1} as its $(k + 1)$-prefix, $C_1^{(j)} C_2^{(j+1)} = A_1 A_2 (B_1 B_2)$ for $j = 1, 2, \ldots, k$ holds for $\sigma = 0(1)$, that is, $C_1^{(j)} C_2^{(j+1)} = [0]$. Hence, we obtain

$$\bar{A}^{(\mathrm{L}+k)}(\sigma^k z) = \bar{A}^{(\mathrm{L})}(z)$$

holds.

THEOREM 6. *Let z' be any tape of length* L, *then*

$$\gamma^{\mathrm{L}} \| H \| \leqslant \| \bar{A}^{(\mathrm{L})}(0z') - \bar{A}^{(\mathrm{L})}(1z') \| \leqslant \delta^{\mathrm{L}} \| H \|,$$

where $\gamma = \min\{\| \lambda_2 \|, \| \mu_2 \|\}$.

This theorem shows that the difference of the two L-th approximate matrices corresponding to two tapes of length $\mathrm{L} + 1$ which have a common L suffix z' is bounded by the upper and lower bounds. It can be shown that this upper or lower bound is satisfied in the case of the tape

$$z' = \overbrace{000 \cdots 0}^{\mathrm{L}} \quad \text{or} \quad \overbrace{11 \cdots 1}^{\mathrm{L}}.$$

Proof. Without loss of generality, we assume that $A(z') = C^{(2)}C^{(3)} \cdots C^{(L)}$ and $C^{(2)} = A$ hold. Then,

$$\bar{A}^{(L)}(1z') = C_1^{(L)} + v_2^{(L-1)}C_1^{(L-1)}C_2^{(L)} + \cdots + v_2^{(3)} \cdots v_2^{(L)}C_1^{(2)}C_2^{(3)} + v_2^{(1)} \cdots v_2^{(L)}B_1C_2^{(2)}, \tag{5}$$

$$\bar{A}^{(L)}(0z') = C_1^{(L)} + v_2^{(L-1)}C_1^{(L-1)}C_1^{(L-1)}C_2^{(L)} + \cdots + v_2^{(3)} \cdots v_2^{(L)}C_1^{(2)}C_2^{(3)} + v_2^{(1)} \cdots v_2^{(L)}A_1C_2^{(2)}. \tag{6}$$

By the assumption $C^{(2)} = A$, we have $A_1C_2^{(1)} = A_1A_2 = [0]$. From (5) and (6), we obtain

$$\| \bar{A}^{(L)}(1z') - \bar{A}^{(L)}(0z')\| = | v_2^{(1)} \cdots v_2^{(L)}| \, \| B_1C_2^{(1)}\| = | v_2^{(1)} \cdots v_2^{(L)}| \, \| H \|. \tag{7}$$

If in (7) we replace $v_2^{(j)}$ by $\gamma(\delta)$, then the left (right) inequality in Theorem 6 is obtained. Q.E.D.

V. Completely Isolated Sets of Stochastic Matrices

Using the properties developed in the previous section, we introduce the new concept of a *completely isolated set* of stochastic matrices, and relate this concept to probabilistic automata. In the rest of this paper, we treat probabilistic automata for which $\delta = \max\{|\lambda_2|, |\mu_2|\} < 1$ holds because Theorem 4 can be applied to this case. We assume that a probabilistic automaton with two states has s_1 as its initial state and s_2 as its final state. Then $p(x)$, the output probability of a tape x being accepted, equals the (1, 2)-element of the matrix $A(x)$. We define the L-th approximate probability $\bar{p}^{(L)}(x)$ as follows.

Definition 7. The L-th *approximate probability* $\bar{p}^{(L)}(x)$ to a tape x is defined to be the (1, 2)-element of the matrix $\bar{A}^{(L)}(x)$.

Corollary 7. *The following relations hold:*

(1) $\bar{p}^{(L)}(\Sigma^* z) = \bar{p}^{(L)}(z)$,
(2) $| p(\Sigma^* z) - \bar{p}^{(L)}(z)| < \bar{\epsilon}^{(L)} = \delta^L/(1-\delta) \max\{\| H \|, \| A_2 \|, \| B_2 \|\}$,
(3) $\bar{p}^{(L)}(z) = \bar{p}^{(L+k+1)}(0^k z)$,
(4) $\gamma^L \| H \| \leqslant | \bar{p}^{(L)}(0z') - \bar{p}^{(L)}(1z')| \leqslant \delta^L \| H \|$,

where a tape z of length $\mathrm{L}+1$ *is written $\sigma z'$, that is, σ is the 1-prefix of Z and z' is the* L*-suffix of z. Also, $\| D \|$ is the absolute value of the* (1, 2)*-element of the matrix D.*

Proof. It is obvious from Definition 6 and 7 that Proposition 2, Theorem 4, 5 and 6 can be rewritten as above. Q.E.D.

DEFINITION 8. Let x be any tape of length greater than or equal to $\mathrm{L}+1$, then an *error interval by the* L-th *approximation* $I_x^{(\mathrm{L})}$ is defined as follows:

$$I_x^{(\mathrm{L})} = [\bar{p}^{(\mathrm{L})}(x) - \bar{\epsilon}^{(\mathrm{L})}, \bar{p}^{(\mathrm{L})}(x) + \bar{\epsilon}^{(\mathrm{L})}].$$

PROPOSITION 3. *Let $p(x)$ and $I_x^{(\mathrm{L})}$ be defined as in Definition 7 and 8, respectively, and let z be a tape of length* $\mathrm{L}+1$, *then*

(1) $I_{\Sigma^* z}^{(\mathrm{L})} = I_z^{(\mathrm{L})}$,

(2) $p(x) \in I_x^{(\mathrm{L})}$, *whenever* $\bar{\epsilon}^{(\mathrm{L})}(x) \neq 0$ *holds.*

Proof. We can easily show that the first property is derived from Proposition 2 and Corollary 7(1). The second property is derived from Corollary 7(2). Q.E.D.

THEOREM 8. *Let x be any tape of length greater than or equal to* $\mathrm{L}+2$, *then*

$$I_x^{(\mathrm{L}+1)} \subset I_x^{(\mathrm{L})}.$$

Proof. Let $A(x)$ be $C^{(1)}C^{(2)} \cdots C^{(m)}$, where $m = \mathrm{L}+2$, then

$$\bar{A}^{(\mathrm{L}+1)}(x) = \bar{A}^{(\mathrm{L})}(x) + v_2^{(m-\mathrm{L})} \cdots v_2^{(m)} C_1^{(m-\mathrm{L}-1)} C_2^{(m-\mathrm{L})}. \tag{8}$$

First, consider the case in which the $(m-\mathrm{L}-1)$-th symbol of x is the same as the $(m-\mathrm{L})$-th symbol of x, then

$$C_1^{(m-\mathrm{L}-1)} C_2^{(m-\mathrm{L})} = [0]$$

holds. Therefore,

$$\bar{A}^{(\mathrm{L}+1)}(x) = \bar{A}^{(\mathrm{L})}(x)$$

holds. Moreover, using definition of $\bar{\epsilon}^{(\mathrm{L})}$, it can be shown that $\bar{\epsilon}^{(\mathrm{L}+1)} < \bar{\epsilon}^{(\mathrm{L})}$ holds.

Second, consider the case in which the $(m - \mathrm{L} - 1)$-th symbol of x is not the same as the $(m - \mathrm{L})$-th symbol of x; then substituting γ and δ for $v_2^{(j)}$ in (8), respectively, we obtain the following inequalities:

$$\gamma^{\mathrm{L}+1} \| H \| \leqslant \| \bar{A}^{(\mathrm{L}+1)}(x) - \bar{A}^{(\mathrm{L})}(x) \| \leqslant \delta^{\mathrm{L}+1} \| H \|,$$

thus,

$$\bar{p}^{(\mathrm{L}+1)}(x) - \delta^{\mathrm{L}+1} \| H \| \leqslant \bar{p}^{(\mathrm{L}+1)}(x) \leqslant \bar{p}^{(\mathrm{L})}(x) + \delta^{\mathrm{L}+1} \| H \|. \tag{9}$$

holds. Then we add $\bar{\epsilon}^{(\mathrm{L}+1)}$ to both side of the second inequality of (9), and we obtain

$$\bar{p}^{(\mathrm{L}+1)}(x) + \bar{\epsilon}^{(\mathrm{L}+1)} \leqslant \bar{p}^{(\mathrm{L})}(x) + \bar{\epsilon}^{(\mathrm{L}+1)} + \delta^{\mathrm{L}+1} \| H \|.$$

From the definition of $\bar{\epsilon}^{(\mathrm{L})}$, it follows that

$$\bar{\epsilon}^{(\mathrm{L}+1)} + \delta^{\mathrm{L}+1} \| H \| < \bar{\epsilon}^{(\mathrm{L})}.$$

Therefore, we obtain

$$\bar{p}^{(\mathrm{L}+1)}(x) + \bar{\epsilon}^{(\mathrm{L}+1)} < \bar{p}^{(\mathrm{L})}(x) + \bar{\epsilon}^{(\mathrm{L})}. \tag{10}$$

Similarly, as in the first inequality of (5.2), the following relation is satisfied,

$$\bar{p}^{(\mathrm{L})}(x) - \bar{\epsilon}^{(\mathrm{L})} < \bar{p}^{(\mathrm{L}+1)}(x) - \bar{\epsilon}^{(\mathrm{L}+1)}. \tag{11}$$

From (10) and (11), it follows that

$$(\bar{p}^{(\mathrm{L}+1)}(x) - \bar{\epsilon}^{(\mathrm{L}+1)}, \bar{p}^{(\mathrm{L}+1)}(x) + \bar{\epsilon}^{(\mathrm{L}+1)}) \subset (\bar{p}^{(\mathrm{L})}(x) - \bar{\epsilon}^{(\mathrm{L})}, \bar{p}^{(\mathrm{L})}(x) + \bar{\epsilon}^{(\mathrm{L})}).$$

Q.E.D.

DEFINITION 9. A pair of tapes x and y of the same length, $\mathrm{L} + 1$, are said to be *isolated by the* L-th *approximation* if and only if $I_x^{(\mathrm{L})} \cap i_y^{(\mathrm{L})} = \varnothing$. A set of two stochastic matrices is said to be *completely isolated by the* L-th *approximation* if and only if every pair of tapes of length $\mathrm{L} + 1$, which is produced from the set of matrices, is isolated by the L-th approximation. In this case, all $2^{\mathrm{L}+1}$ intervals $I_{\Sigma * z}^{(\mathrm{L})}$, corresponding to a tape $\Sigma^* z$ of length greater than or equal to $\mathrm{L} + 1$, have no common point in the unit interval [0, 1].

LEMMA 2. *A necessary and sufficient condition for the set of matrices being completely isolated by the* 0-th *approximation, is*

$$\| H \| \neq 0 \qquad \textit{and} \qquad h \leqslant 1,$$

where

$$h = \frac{2\delta}{1-\delta} \frac{\max\{\| H \|, \| A_2 \|, \| B_2 \|\}}{\| H \|}.$$

Proof. From Table II,

$$\bar{A}^{(0)}(0) = A_1 = \begin{bmatrix} b/(a+b) & a/(a+b) \\ b/(a+b) & a/(a+b) \end{bmatrix},$$

$$\bar{A}^{(0)}(1) = B_1 = \begin{bmatrix} d/(c+d) & c/(c+d) \\ d/(c+d) & c/(c+d) \end{bmatrix},$$

and by Definition 6,

$$\bar{\epsilon}^{(0)} = \frac{\delta}{1-\delta} \max\{\| H \|, \| A_2 \|, \| B_2 \|\},$$

therefore, two error intervals by the L-th approximation $I_0^{(0)}$ and $I_1^{(0)}$ are isolated if and only if

$$| \bar{p}^{(0)}(0) - \bar{p}^{(0)}(1) | \geqslant 2\bar{\epsilon}^{(0)}. \tag{12}$$

Now the (1, 2)-element is designated, so the leftside of (12) can be rewritten as

$$| a/(a+b) - c/(c+d) | = |(ad-bc)/(a+b)(c+d)| = \| H \|.$$

Therefore, (12) implies

$$\| H \| \geqslant 2\bar{\epsilon}^{(0)} = 2 \frac{\delta}{1-\delta} \max\{\| H \|, \| A_2 \|, \| B_2 \|\}. \tag{13}$$

Hence, we obtain

$$h \leqslant 1. \qquad \text{Q.E.D.}$$

We know from (13) that if $\| H \| = 0$, then $\bar{\epsilon}^{(0)} = 0$. This implies that both $I_0^{(0)}$ and $I_1^{(0)}$ are void intervals and this case should be omitted.

THEOREM 9. *Assume that a given set of two stochastic matrices is isolated by the* 0-th *approximation. Then the condition*

$$h\delta^{\mathrm{L}} \leqslant \gamma^{\mathrm{L}}$$

is necessary and sufficient for the set being isolated by the L-th *approximation.*

Proof. Using the assumtion that a given set is completely isolated by the 0-th approximation and Theorem 8 and Corollary 7, it follows that the condition

$$\gamma^{\mathrm{L}} \| H \| \geqslant 2\bar{\epsilon}^{(\mathrm{L})} = 2 \frac{\delta^{\mathrm{L}+1}}{1-\delta} \max\{\| H \|, \| A_2 \|, \| B_2 \|\} \tag{14}$$

is necessary and sufficient for the set being completely isolated by the L-th approximation. Using h as defined in Lemma 2, we can rewrite the condition (14) as follows:

$$h\delta^{\mathrm{L}} \leqslant \gamma^{\mathrm{L}}. \qquad \text{Q.E.D.}$$

LEMMA 3. *Suppose that a given set of two stochastic matrices is completely isolated by the* 0-th *approximation but not by the* L-th *approximation. Then for an arbitrary integer* $k(\geqslant 1)$, *the set is not completely isolated by the* $(\mathrm{L}+k)$-th *approximation.*

Proof. By the assumption,

$$h\delta^{\mathrm{L}} > \gamma^{\mathrm{L}}$$

holds. Since $\gamma \leqslant \delta$ must hold, it follows that

$$h\delta^{\mathrm{L}+k} > \gamma^{\mathrm{L}+k}.$$

By Theorem 9, this means the set is not completely isolated by the $(\mathrm{L}+k)$-th approximation. Q.E.D.

LEMMA 4. *If a set is not completely isolated by the* 0-th *approximation, then the set is not completely isolated by the* L-th *approximation for each positive integer* L.

Proof. By the assumption,

$$h > 1.$$

By using $\gamma \leqslant \delta$, we obtain

$$h\delta^{\mathrm{L}} > \gamma^{\mathrm{L}}.$$

Therefore,

$$\gamma^{\mathrm{L}} \| H \| < \frac{2\delta^{\mathrm{L}+1}}{1-\delta} \max\{\| H \|, \| A_2 \|, \| B_2 \|\} = 2\bar{\epsilon}^{(\mathrm{L})}.$$

From Corollary 7, there exists a pair of tapes x and y such that

$$| \bar{p}^{(\mathrm{L})}(x) - \bar{p}^{(\mathrm{L})}(y)| = \gamma^{\mathrm{L}} \| H \|$$

holds. Therefore,

$$| \bar{p}^{(\mathrm{L})}(x) - \bar{p}^{(\mathrm{L})}(y)| < 2\bar{\epsilon}^{(\mathrm{L})}$$

holds. Hence, we get

$$I_x^{(\mathrm{L})} \cap I_y^{(\mathrm{L})} \neq \phi. \qquad \text{Q.E.D.}$$

THEOREM 10. *If a set is completely isolated by the* 0-th *approximation and* $\gamma = \delta$, (*that is, if* $|\lambda_2| = |\mu_2|$), *then for each positive integer* L, *the set is completely isolated by the* L-th *approximation.*

Proof. Using the assumption $h \leqslant 1$ and $\gamma = \delta$, we conclude that for each positive integer L,

$$h\delta^{\mathrm{L}} \geqslant \gamma^{\mathrm{L}}$$

holds. Applying Theorem 10 to the inequality, we conclude that the set is completely isolated by the L-th approximation.

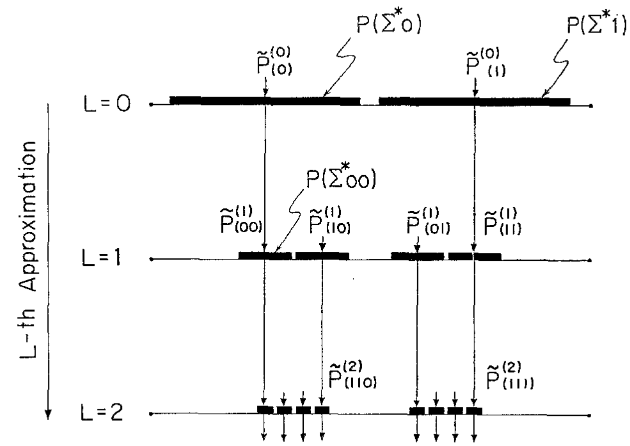

FIG. 3(a). Aspects of the L-th approximate probability $\tilde{P}^{(L)}(x)$ of completely isolated automaton—associated with output probability $P(x)$.

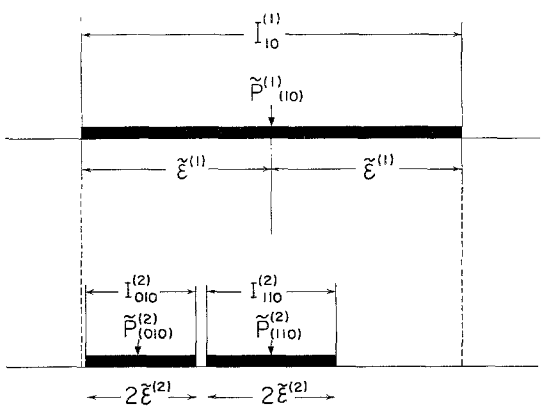

FIG. 3(b). Error intervals $I_{10}^{(1)}$, $I_{010}^{(2)}$, and $I_{110}^{(2)}$—partially enlarged from Fig. 3(a).

In Fig. 3, the behavior of the L-th approximate probabilities $\bar{p}^{(\mathrm{L})}(x)$ of a completely isolated automaton is shown. A completely isolated automaton is easily shown to be nondense and ϵ approximable (Paz [4]).

VI. Capability of Two-State Two-Symbol Probabilistic Automata

In the last part of this paper, we associate completely isolated sets of two stochastic matrices with events realized by probabilistic automata.

Definition 10. A probabilistic automaton is said to be *completely isolated by the* L-th *approximation* if its set of stochastic matrices is completely isolated by the L-th approximation.

Theorem 11. *If a probabilistic automaton PA has a set of stochastic matrices which is completely isolated by the* L-th *approximation, then*

$$T(PA, \bar{p}^{(\mathrm{L})}(z), \bar{\epsilon}^{(\mathrm{L})}) = \Sigma^* z \cup Q,$$

where z is a tape of length L + 1 *and Q is a finte set of tapes of legth less than* L + 1. *Thus, the probabilistic automaton PA realizes an* (L + 1)*-definite event.*

Proof. It is obvious from Corollary 7(2) and Definition 8. Q.E.D.

Example 2. We now consider two examples of probabilistic automata which are completely isolated by the L-th approximation, where L is any positive integer.

TABLE III

Examples of a Completely Isolated Set of Two Stochastic Matrices

(a)

$$A = \begin{bmatrix} 1 & 0 \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix} \qquad B = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ 0 & 1 \end{bmatrix}$$

(b)

$$A = \begin{bmatrix} 0.8 & 0.2 \\ 0.7 & 0.3 \end{bmatrix} \qquad B = \begin{bmatrix} 0.6 & 0.4 \\ 0.5 & 0.5 \end{bmatrix}$$

Let $PA = (S, M, s_1, F)$ be a probabilistic automaton over $\Sigma = \{0, 1\}$ such that $S = \{s_1, s_2\}$, $A(0) = A$, $A(1) = B$, and $F = \{s_2\}$, where the matrices A and B are given in Table III. The probabilistic automaton of

Table III(a) from Rabin [5] is known to be a nondense 3-adic PA. By a simple calculation, we obtain

$$\gamma = \delta = \frac{1}{3}, \qquad \| H \| = \frac{2}{3}, \qquad \| A_2 \| = 0, \qquad \text{and} \qquad \| B_2 \| = 1,$$

therefore,

$$h = 1.$$

Similarly, in the case of Table III(b), we get

$$\gamma = \delta = \frac{1}{10}, \qquad \| H \| = \frac{2}{9}, \qquad \| A_2 \| = \frac{2}{9}, \qquad \text{and} \qquad \| B_2 \| = \frac{4}{9}.$$

Therefore,

$$h = \frac{4}{81} < 1$$

holds. Hence, the set of A and B is completely isolated by the 0-th approximation. Moreover, $\gamma = \delta$ is true. Using these facts and Theorem 4, we conclude that both PA's mentioned above are completely isolated by the L-th approximation. This leads us to the next theorem.

THEOREM 12. *For any positive integer* L, *there exist two-state two-symbol probabilistic automata which can realize any* L-*definite event in the sense of Definition* 4.

Furthermore, using the method given above, we can determine whether or not a given two-state two-symbol probabilistic automaton is completely isolated.

VII. CONCLUDING REMARKS

Two-state two-symbol probabilistic automata are discussed from the viewpoint of characteristic values of the stochastic matrices of input symbols. Although it is difficult to treat general multistate probabilistic automaton in similar manner, certain special multistate probabilistic automata have been discussed in the authors' preceding paper (Yasui and Yajima [8]).

The concept of completely isolated automata introduced here in the extension of 3-adic nondense probabilistic automata in Rabin [5]. We can also explain this concept in a physical sense as follows. Assume that we are given a probabilistic automaton which is completely isolated by the L-th approximation. When an unknown tape x of length greater than L is fed into it, we can approximate $p(x)$ experimentally, although we cannot determine $p(x)$ exactly without infinite trials. Once the estimated $p(x)$ falls into one of the error intervals by the L-th approximation, say $I_y^{(L)}$, then $x = zy$ must hold for some tape z where y is a tape of length $L + 1$. In other words, we can determine the $(L + 1)$-suffix y of the unknown tape x. In this way, all tapes of length greater than or equal to $L + 1$ can be classified into 2^{L+1} sets by means of their $(L + 1)$-suffixes.

While we were developing this algebraic treatment of probabilistic automata with two states, Paz [4] developed many interesting concepts and results. He shows that probabilistic automata with two states are ϵ approximable. If we use this concept, probabilistic automata which are completely isolated by the L-th approximation can be said to be $\epsilon^{(L)}$ approximable by deterministic definite automata with 2^{L+1} states, which correspond to 2^{L+1} output probabilities $p^{(L)}(\Sigma^* x)$.

Acknowledgments

The authors thank Prof. K. Maeda and Prof. T. Kiyono of Kyoto University for their encouragement and support of this study, and is also grateful for the constructive discussions with their colleagues, especially Mr. Y. Kambayashi and Mr. Y. Yoshida.

They would like to express their appreciation to the referee for his comments and suggestions, especially for his proof in Remark with respect to Definition 4.

Received: August 23, 1968; revised: November 28, 1969

References

1. W. Feller, "An Introduction to Probability Theory and Its Applications," Vol. 1, Wiley, New York, 1958.
2. A. Paz, Definite and quasidefinite sets of stochastic matrices, *Proc. Amer. Math. Soc.* **16** (1965), 643–645.
3. A. Paz, Some aspects of probabilistic automata, *Information and Control* **9** (1966), 26–90.
4. A. Paz, Fuzzy star functions, probabilistic automata and their approximation by nonprobabilistic automata, Conf. Rec. IEEE 8-th Ann. Symp. on Switching and Automata Theory (1967), 280–290.
5. M. O. Rabin, Probabilistic automata, *Information and Control* **6** (1963), 230–245.
6. A. Salomaa, On *m*-adic probabilistic automata, *Information and Control* **10** (1967), 215–219.
7. P. Turakainen (1968), On probabilistic automata and their generalizations, *Ann. Acad. Sci. Fenn.*, Series A.I. Math. vol. 429.
8. T. Yasui and S. Yajima, Some algebraic properties of sets of stochastic matrices, *Information and Control* **14** (1969), 319–357.